Luc Van Gool (Ed.), Pattern Recognition,
Zürich, Sept. 16–18, 2002, LNCS Vol. 2449, 472–480, © Springer.

Motion Competition: Variational Integration of Motion Segmentation and Shape Regularization[1]

Daniel Cremers and Christoph Schnörr

Computer Vision, Graphics and Pattern Recognition Group
Department of Mathematics and Computer Science
University of Mannheim, D–68131 Mannheim, Germany
{cremers,schnoerr}@uni-mannheim.de
http://www.cvgpr.uni-mannheim.de

Abstract. We present a variational method for the segmentation of piecewise affine flow fields. Compared to other approaches to motion segmentation, we minimize a *single* energy functional both with respect to the affine motion models in the separate regions and with respect to the shape of the separating contour. In the manner of region competition, the evolution of the segmenting contour is driven by a force which aims at maximizing a homogeneity measure with respect to the estimated motion in the adjoining regions.
We compare segmentations obtained for the models of piecewise affine motion, piecewise constant motion, and piecewise constant intensity. For objects which cannot be discriminated from the background by their appearance, the desired motion segmentation is obtained, although the corresponding segmentation based on image intensities fails. The region–based formulation facilitates convergence of the contour from its initialization over fairly large distances, and the estimated discontinuous flow field is progressively improved during the gradient descent minimization. By including in the variational method a statistical shape prior, the contour evolution is restricted to a subspace of familiar shapes, such that a robust estimation of irregularly moving shapes becomes feasible.

Keywords: Region Competition, Motion segmentation, piecewise affine motion, variational methods, statistical shape prior.

1 Related Work

Discontinuity–preserving motion estimation by variational methods and related partial differential equations have a long tradition in computer vision. In some approaches the motion discontinuities are modeled implicitly in terms of appropriate (non–quadratic) regularizers [14, 2, 12, 11, 17]. Other approaches pursue separate steps of variational motion estimation on disjoint sets with a shape optimization procedure [16, 4, 15, 8]. For the case of grey value segmentation, there exist some region–based variational approaches with explicit discontinuities (cf. [13]) and extensions to color and texture segmentation [18].

In this paper, we present a variational method for motion segmentation with an explicit contour description. The problems of segmentation and motion estimation are jointly solved by gradient descent on a *single* energy functional. In

[1] This contribution was awarded the main prize of the conference.

contrast to implicit level set based shape representations (cf. [4]), the *explicit* representation of the contour does not permit topological changes. However, it facilitates the incorporation of a statistical prior on the shape of the segmenting contour. Furthermore, in many applications it is known that topological changes of shapes do not occur, in which case a fixed topology is preferable.

Deformable shape models are combined with motion segmentation in [10]. However, there the authors do not propose a variational integration of motion segmentation and shape prior. Rather they optimize a small number of shape parameters by simulated annealing, which — unlike our approach — cannot be applied to more general shape priors.

2 Variational Motion Segmentation

Let $f(x,t)$ be an image sequence which is assumed to be differentiable. If the intensity of a moving point is constant throughout time, we obtain a continuity equation given by the classical optic flow constraint:

$$\frac{d}{dt} f(x,t) = \frac{\partial}{\partial t} f + w^t \nabla f = 0,$$

where $w = \frac{dx}{dt}$ denotes the local velocity. Given two consecutive images f_1 and f_2 from this sequence, we can approximate $\frac{\partial}{\partial t} f \approx (f_2 - f_1)$ and $\nabla f \approx \frac{1}{2}\nabla(f_1 + f_2)$.

We propose to segment the image plane into areas R_i of parametric motion $w_i = w(\alpha_i)$ by minimizing the energy functional

$$E(\alpha, C) = \sum_i \int\limits_{R_i} \left(f_2 - f_1 + \frac{w_i^t}{2} \nabla(f_1 + f_2) \right)^2 dx + \nu\, E_c(C) \tag{1}$$

simultaneously with respect to both the contour C which separates the regions R_i, and the motion parameters $\alpha = \{\alpha_i\}$. The term E_c represents an *internal shape energy*, such as the length of the contour or a more elaborate shape dissimilarity measure, which will be detailed in Section 5.

With the extended velocity vector $v = \binom{w}{1}$ and the spatio–temporal structure tensor [1]

$$S = (\nabla_3 f)(\nabla_3 f)^t, \quad \text{with } \nabla_3 f = \begin{pmatrix} \nabla f \\ \frac{\partial}{\partial t} f \end{pmatrix},$$

the energy (1) can be rewritten as

$$E(\alpha, C) = \sum_i \int\limits_{R_i} \left(v_i^t\, S\, v_i \right) dx \;+\; \nu\, E_c(C). \tag{2}$$

In practice, the homogeneity term shows a bias towards velocity vectors of large magnitude. As proposed in [7], we therefore perform an isotropy compensation of the structure tensor by replacing S with $S - \lambda_3\, I$, where λ_3 is the smallest eigenvalue of S and I is the 3×3 unit matrix.

3 Piecewise Homogeneous Motion

The proposed *motion energy* (2) can be interpreted as an extension of the Mumford–Shah model [13, 6] to the problem of motion segmentation. Rather than measuring the grey value homogeneity, it measures the homogeneity with respect to parametric motion models in the respective regions. In the following we will focus on the two cases of constant motion and affine motion, but depending on the application other motion models can be used, as long as the extended velocity vector is linear in the parameters.

For the model of **piecewise constant motion**, the extended velocity vector for region R_i is given by:

$$v_i = T\alpha_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} (a_i\ b_i\ 1)^t, \tag{3}$$

where a_i and b_i denote the velocity in x– and y–direction. For the model of **piecewise affine motion**, we have:

$$v_i = T\alpha_i = \begin{pmatrix} x & y & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x & y & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} (a_i\ b_i\ c_i\ d_i\ e_i\ f_i\ 1)^t, \tag{4}$$

with 6 parameters defining the motion in region R_i.

Inserting the respective parametric motion model into the motion energy (2), we get:

$$E(\alpha, C) = \sum_i \alpha_i^t\, Q_i\, \alpha_i \ + \ \nu\, E_c(C), \tag{5}$$

where

$$Q_i = \int_{R_i} T^t S T \, dx = \begin{pmatrix} \bar{Q}_i & q_i \\ q_i^t & \gamma_i \end{pmatrix}.$$

Depending on the model, the submatrix $\bar{Q}_i$ and the vector q_i have the dimension 2 for the constant motion model or 6 for the affine model.

4 Energy Minimization

The motion energy (5) must be simultaneously minimized both with respect to the evolving contour and with respect to the motion parameters $\{\bar{\alpha}_i\}$, where $\alpha_i = \binom{\bar{\alpha}_i}{1}$ is defined with respect to the chosen motion model — see equations (3) and (4).

Minimization with respect to $\bar{\alpha}_i$ results in the linear equation $\bar{Q}_i\, \bar{\alpha}_i = -q_i$. Due to the well–known *aperture problem*, the symmetric square matrix $\bar{Q}_i$ may not be invertible. In this case, we impose an additional constraint by choosing the solution of minimal length. This amounts to applying the pseudo–inverse (cf. [7]):

$$\bar{\alpha}_i = -\bar{Q}_i^{\dagger}\, q_i. \tag{6}$$

Using Green's theorem (cf. [18]), minimization of (2) with respect to the contour C results in the gradient descent evolution equation

$$\frac{dC}{d\tau} = -\frac{dE}{dC} = \left(e^- - e^+\right) n - \nu \frac{dE_c}{dC} \tag{7}$$

The last term minimizes the internal shape energy which will be treated in the next section. The superscripts $j = +/-$ denote the two regions to the left and to the right of the respective contour point (in the sense of the contour parameterization), and n is the normal on the contour pointing out of the region R_+.

The adjacent regions compete for the contour in terms of the associated energy densities[2]

$$e^j = v_j^t \, S \, v_j. \tag{8}$$

This *motion competition* enforces regions of homogeneous optic flow, thus separating regions moving at different velocities w_j.

Following the argumentation in [7], we normalize the cost function in (8) by replacing $v_j^t \, S \, v_j$ with $\frac{v_j^t \, S \, v_j}{||v_j||^2 \, trS}$. Although this modification is not strictly derived by minimizing energy (5), it tends to slightly improve the contour evolution.

5 Internal Shape Energy

In the following, we will present two possible models for the internal shape energy E_c in (5). We will restrict the space of possible motion contours to closed spline curves of the form $C : [0,1] \to \Omega, \quad C(s) = \sum_{n=1}^{N} p_n \, B_n(s)$, with spline control points $p_n = (x_n \, , \, y_n)^t$ and periodic quadratic basis functions B_n [3]. This permits a relatively fast numerical optimization. Moreover, it facilitates the incorporation of a statistical shape prior on the control point vector $z = (x_1, y_1, \ldots, x_N, y_N)^t$.

The first and fairly general internal energy is given by a measure of the contour length:

$$E_c(C) := \frac{1}{2} \int_0^1 \left(\frac{dC}{ds} \right)^2 ds. \tag{9}$$

In the case of quadratic B–spline basis functions, the corresponding Euler–Lagrange equation is equivalent to an equidistant spacing of control points. The term $-\nu \frac{dE_c}{dC}$ in the evolution equation (7) simply pulls each control point towards the center of its respective neighbors.

For the second choice of internal energy, we construct a *shape dissimilarity measure* which encodes statistically the silhouettes of a set of sample shapes [6]. To this end, the images of training objects are binarized, a spline contour is

[2] In the equivalent probabilistic interpretation, this energy density represents the *log likelihood* for the probability that a given location is part of one or the other motion hypothesis.

fitted to the boundary, and the set of training contours is aligned with respect to similarity transformations [9] and cyclic permutation of the control points.

The distribution of control point vectors $z \in \mathbb{R}^{2N}$ is assumed to be Gaussian: $\mathcal{P}(z) \propto \exp\left(-\frac{1}{2}(z-z_0)^t \, \Sigma^{-1} \, (z-z_0)\right)$. The mean control point vector z_0 and sample covariance matrix Σ are determined for the training set.[3] The negative logarithm of the Gaussian probability can be interpreted as a shape energy of Mahalanobis type:

$$E_{shape}(z) = \frac{1}{2} \, (z-z_0)^t \, \Sigma^{-1} \, (z-z_0). \tag{10}$$

By construction, this energy is not invariant with respect to transformations such as translation or rotation of the shape. As shown in [5], there is a closed–form solution for incorporating such invariances in the variational approach. One simply applies the function (10) to the argument after alignment of the respective contour with respect to the mean shape z_0:

$$E_c(z) := E_{shape}\left(\frac{R(z-z_c)}{|R(z-z_c)|}\right), \tag{11}$$

where z_c denotes the centered version of z and R denotes the optimal rotation with respect to the mean z_0. The resulting expression can be differentiated with respect to the control point vector z (cf. [5]). This incorporates similarity invariance on the basis of the control point polygons without any additional parameters to encode rotation angle, scale and translation.

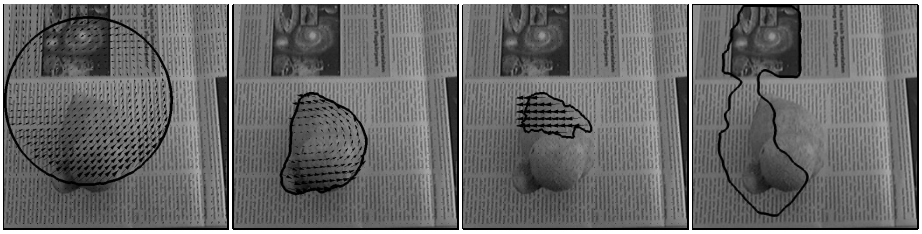
6 Evolution of the Motion Boundary

Equation (7) can be converted to an evolution equation for the spline control points by inserting the spline representation of the contour. The equation is discretized with a set of nodes s_i along the contour, where s_i is chosen as the point where the respective spline basis function B_i attains its maximum. Including the contribution of the internal shape energy, we obtain for the x–coordinate of control point m:

$$\frac{dx_m(t)}{d\tau} = \sum_k \left(\mathbf{B}^{-1}\right)_{mk} \left(e^+_{s_k} - e^-_{s_k}\right) n_x(s_k) \; - \; \nu \left(\frac{dE_c}{dz}\right)_{2m-1}, \tag{12}$$

where n_x denotes the x–coordinate of the normal vector and the index $2m-1$ refers to the component of the given vector which is associated with the x–coordinate of control point m. The cyclic tridiagonal matrix $\mathbf{B}$ contains the spline basis functions evaluated at the nodes: $B_{ij} = B_i(s_j)$. A similar expression holds for the y–coordinates.

[3] If the dimension of the subspace spanned by the training vectors is smaller than the dimension $2N$ of the underlying vector space, the sample covariance matrix Σ is regularized by replacing the non–zero eigenvalues by a constant $\sigma_\perp = 0.5 \, \sigma_r$, where σ_r is the smallest non–vanishing eigenvalue. For a justification we refer to [6].



Initial (affine) Affine motion Constant motion Constant intensity

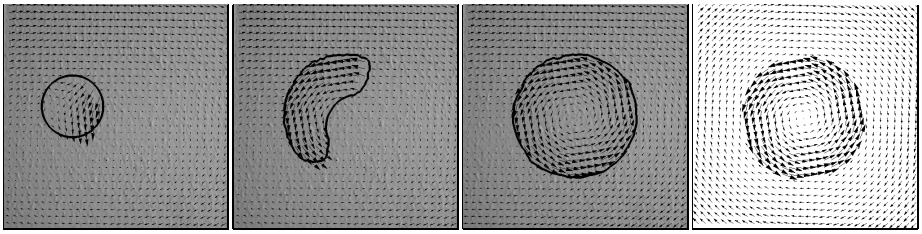
Fig. 1. Model comparison: Initial contour and final segmentation obtained by minimizing the functional (5) with the piecewise affine motion model (4), and final segmentations (with the same initialization) for the model (3) of piecewise constant motion, and the corresponding model of piecewise constant intensity [6]. The input images show a duck figure rotated on a newspaper. The affine motion model captures the rotation and thereby correctly segments the duck. The model of constant motion only captures those parts which show approximately constant motion, whereas the model of constant intensity is mislead by background clutter and the difficult lighting conditions.

The two terms in (12) can be interpreted as follows: The first term forces the contour towards the boundaries of the two motion fields by minimizing the motion inhomogeneity in the adjoining regions, measured by the energy density (8). The last term minimizes the internal shape energy — in our case the length of the contour (9), a shape dissimilarity measure of the form (11) or a linear combination of both. The total energy (5) is minimized by iterating the contour evolution (12) in alternation with the update (6) of the motion estimates.

7 Experimental Results

Model Comparison. Figure 1 shows a comparison of segmentation results obtained for the model (5) of **piecewise affine motion** (4), **piecewise constant motion** (3) and the corresponding model of **piecewise constant intensity** [6]. With the affine model the object is correctly segmented and the rotational motion is correctly estimated. In particular, the estimated center of rotation converges towards the correct one (not shown here) during the minimization. The model of piecewise constant motion only segments a small region of the object which complies with the hypothesis of constant motion. The last image shows that the hypothesis of constant intensity (i.e. average grey value) is obviously not applicable here due to clutter and difficult lighting conditions.

Moving Background. Figure 2 shows an example of two differently moving regions. We took a snapshot of a section of wallpaper and artificially rotated a circular region in the center, the remaining area was rotated in the opposite sense (see ground truth on the right). The contour evolution shows how the two affine motion fields are progressively separated during the energy minimization. The final segmentation cannot be visually distinguished from the rotated circular area.



Contour evolution for the model of piecewise affine motion True motion field

Fig. 2. Separating two affine motion fields: Gradient descent evolution for the functional (5) with the piecewise affine motion model (4). The input sequence shows a wallpaper with a circular area in the center rotated in one sense, and the background rotated in the opposite sense, as indicated by the true motion field on the right. During energy minimization the estimated motion fields are continuously improved, the two motion fields are separated, and the circular area is correctly segmented. Note that the circular area cannot be detected based on grey value information.

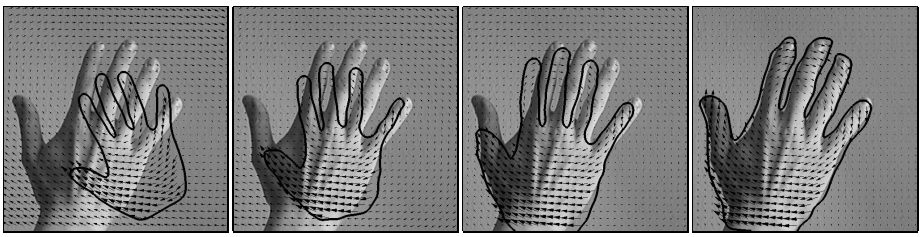


Fig. 3. Knowledge–driven motion segmentation: Gradient descent on the functional (5) with the piecewise affine motion model (4) and a statistical prior (11) favoring hand shapes. The input sequence is a rotating hand in front of a static background. During the contour evolution the estimated motion fields are continuously improved, while the statistical prior restricts the contour to the submanifold of familiar shapes.

Motion Segmentation with a Statistical Shape Prior. Figure 3 shows a contour evolution with a statistical prior favoring hands which was automatically generated from a set of 10 binary training images [6]. The evolving contour is effectively restricted to the submanifold of familiar shapes. While the two estimated affine flow fields are initially fairly similar, they are progressively improved during the energy minimization. In particular, the rotatory motion of the hand and the static background are correctly determined in the final segmentation.

8 Conclusion

We presented a variational method for the segmentation of piecewise affine flow fields. In the manner of region competition, the evolution of the segmenting contour is driven by a force which aims at maximizing a homogeneity measure with respect to the estimated motion in the adjoining regions. By minimizing a *single* functional, we jointly solve the problems of segmentation and motion estimation. Experimental results demonstrate that objects which are not discernible

by their appearance can be segmented according to their relative motion. The region–based formulation permits a convergence of the contour over large distances. Moreover, by the variational integration of a *statistical shape prior*, the evolving contour can be effectively restricted to a subspace of familiar shapes, which facilitates the robust segmentation of more complex moving shapes.

References

1. J. Bigün, G. H. Granlund, and J. Wiklund. Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 13(8):775–790, 1991.
2. M. J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise–smooth flow fields. *Comp. Vis. Graph. Image Proc.: IU*, 63(1):75–104, 1996.
3. A. Blake and M. Isard. *Active Contours*. Springer, London, 1998.
4. V. Caselles and B. Coll. Snakes in movement. *SIAM J. Numer. Anal.*, 33:2445–2456, 1996.
5. D. Cremers, T. Kohlberger, and C. Schnörr. Nonlinear shape statistics in Mumford–Shah based segmentation. In A. Heyden et al., editors, *Proc. of the Europ. Conf. on Comp. Vis.*, volume 2351 of *LNCS*, pages 93–108, Copenhagen, May, 28–31 2002. Springer, Berlin.
6. D. Cremers, F. Tischhäuser, J. Weickert, and C. Schnörr. Diffusion–snakes: Introducing statistical shape knowledge into the Mumford–Shah functional. *Int. J. of Comp. Vis.* Accepted for publication.
7. G. Farnebäck. *Spatial Domain Methods for Orientation and Velocity Estimation*. PhD thesis, Dept. of Electrical Engineering, Linköpings universitet, 1999.
8. G. Farnebäck. Very high accuracy velocity estimation using orientation tensors, parametric motion, and segmentation of the motion field. In *Proc. 8th Int. Conf. on Computer Vision*, volume 1, pages 171–177, 2001.
9. C. Goodall. Procrustes methods in the statistical analysis of shape. *J. Roy. Statist. Soc., Ser. B.*, 53(2):285–339, 1991.
10. C. Kervrann and F. Heitz. Statistical deformable model-based segmentation of image motion. *IEEE Trans. on Image Processing*, 8:583–588, 1999.
11. P. Kornprobst, R. Dériche, and G. Aubert. Image sequence analysis via partial differential equations. *J. Math. Imag. Vision*, 11(1):5–26, 1999.
12. E. Mémin and P. Pérez. Dense estimation and object–based segmentation of the optical flow with robust techniques. *IEEE Trans. on Im. Proc.*, 7(5):703–719, 1998.
13. D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.
14. P. Nesi. Variational approach to optical flow estimation managing discontinuities. *Image and Vis. Comp.*, 11(7):419–439, 1993.
15. J.-M. Odobez and P. Bouthemy. Direct incremental model–based image motion segmentation for video analysis. *Signal Proc.*, 66:143–155, 1998.
16. C. Schnörr. Computation of discontinuous optical flow by domain decomposition and shape optimization. *Int. J. of Comp. Vis.*, 8(2):153–165, 1992.
17. J. Weickert and C. Schnörr. A theoretical framework for convex regularizers in PDE–based computation of image motion. *Int. J. of Comp. Vis.*, 45(3):245–264, 2001.
18. S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 18(9):884–900, 1996.